**Individual Project**

**Sakshi Sandeep Salvi**

**Business Analytics, Conestoga College, Institute of Technology and Advanced Learning**

**INFO8076: SQL and Data Analysis**

**Jey Kumaresan**

**Feb 23, 2025**

1. **Goal of the project – What do you want to learn from this data?**

The dataset examined contains information about contact details for public schools in Ontario. The research project will look at public educational data that covers school board information and all available academic institutions across Ontario. The research analyzes educational institution relationships through their classification as well as their geographical positions and their administrative systems. This study allows us to identify patterns about school location distribution across regions together with preferred languages and additional elements that affect educational institutions. The database method using SQL retrieves essential educational data in an efficient manner before its arrangement enables decisions regarding resource distribution and planning.

1. **Data Source and Data Cleaning Process**

**Data Source:** [**https://data.ontario.ca/dataset/ontario-public-school-contact-information**](https://data.ontario.ca/dataset/ontario-public-school-contact-information)

**Data Cleaning Process:**

1. Duplicate elimination ensured each school received the same record.
2. The process used uniform column names for consistent datasets.
3. Every null field received default values through assignment using empty value handlers.
4. The database contains complete specifications for converting date values, numeric data, and text data types.
5. A process of dataset optimization required the removal of unneeded columns.

The standard data preparation architecture produced precise data suitable for performing SQL analytical operations easily.

1. **Relational schema/ERD**

A screenshot of a computer

Description automatically generated

1. **SQL Queries**

|  |
| --- |
| **Query 1:** Retrieve distinct school languages  SELECT DISTINCT SCHOOL\_LANGUAGE  FROM ADDITIONAL\_DETAILS; |
| **Output** |
| **Insight**  This question aids in determining the range of languages spoken by the schools in various regions. The institutions mostly offer academic programs in either French or English. |

|  |
| --- |
| **Query 2:** Retrieve all high schools sorted by name descending  SELECT SCHOOL\_NAME  FROM SCHOOL\_DETAILS  WHERE SCHOOL\_LEVEL = 'Secondary'  ORDER BY SCHOOL\_NAME DESC; |
| **Output** |
| **Insight**  Based on alphabetical order, the query generates a descending sort of secondary schools. In addition to helping with naming system analysis, the system makes it simple for users to locate schools. A large number of secondary schools indicates that higher education institutions are distributed appropriately. Here, there are a total of 30 secondary level schools. |

|  |
| --- |
| **Query 3:** Find all schools that are in a specific region (e.g., 'North')  SELECT s.SCHOOL\_NAME, b.REGION\_NAME  FROM SCHOOL\_DETAILS s  INNER JOIN BOARD\_DETAILS b ON s.BOARD\_ID = b.BOARD\_ID  WHERE b.REGION\_NAME = 'North Region'; |
| **Output** |
| **Insight**  The query makes school filtration possible, which separates educational institutions geographically. Because it displays the number of educational institutions serving certain zones and their resource allocation patterns across regions, the query provides crucial insights on the distribution of schools by region. In the North Region, there are 19 schools. |

|  |
| --- |
| **Query 4:** Find schools that are neither Primary nor Secondary level  SELECT SCHOOL\_NAME, SCHOOL\_LEVEL  FROM SCHOOL\_DETAILS  WHERE SCHOOL\_LEVEL NOT IN ('Secondary'); |
| **Output** |
| **Insight**  The analysis identifies particular educational establishments, such as elementary, secondary or elementary and secondary institutions. In this case, 40 schools offer more than just secondary education; the other schools offer either elementary education or both elementary and secondary education. |

|  |
| --- |
| **Query 5:** Find the number of schools per school type  SELECT SCHOOL\_TYPE, COUNT(SCHOOL\_TYPE) AS TOTAL\_SCHOOLS  FROM ADDITIONAL\_DETAILS  GROUP BY SCHOOL\_TYPE; |
| **Output** |
| **Insight**  A summary of the distribution of various educational institutions (Catholic, public, hospital, and many more) within the population is provided by the query. The majority of the schools in this region are either public or Catholic. |

|  |
| --- |
| **Query 6:** Find the number of schools under each board, but only show boards with more than 2 schools  SELECT b.BOARD\_NAME, COUNT(s.SCHOOL\_ID) AS SCHOOL\_COUNT  FROM SCHOOL\_DETAILS s  INNER JOIN BOARD\_DETAILS b ON s.BOARD\_ID = b.BOARD\_ID  GROUP BY b.BOARD\_NAME  HAVING COUNT(s.SCHOOL\_ID) >= 2; |
| **Output** |
| **Insight**  Multiple-school governing school boards are chosen by the query, but boards with few schools are not included in the results. In this case, the Peel DSB board is in charge of two schools. |

|  |
| --- |
| **Query 7:** --Categorizing Schools  SELECT SCHOOL\_NAME, SCHOOL\_LEVEL,  CASE  WHEN SCHOOL\_LEVEL = 'Elementary' THEN 'Primary'  WHEN SCHOOL\_LEVEL = 'Secondary' THEN 'High School'  ELSE 'Other'  END AS SCHOOL\_CATEGORY  FROM SCHOOL\_DETAILS; |
| **Output** |
| **Insight**  Through a consistent system of description, standardized school levels offer basic analysis tools. The organization standard makes it possible to compare primary and secondary education institutions through the reporting procedure and school grouping. |

|  |
| --- |
| **Query 8:** Create a View to Count Schools per Board  CREATE VIEW School\_Count\_By\_Board AS  SELECT  b.BOARD\_NAME,  b.REGION\_NAME,  COUNT(s.SCHOOL\_ID) AS TOTAL\_SCHOOLS  FROM BOARD\_DETAILS b  LEFT JOIN SCHOOL\_DETAILS s ON b.BOARD\_ID = s.BOARD\_ID  GROUP BY b.BOARD\_NAME, b.REGION\_NAME  ORDER BY TOTAL\_SCHOOLS DESC;  SELECT \*  FROM School\_Count\_By\_Board; |
| **Output** |
| **Insight**  This view's precomputed summary displays the distribution of schools across several boards. For issues pertaining to resource allocation and judgments on educational policy, the view allows stakeholders to quickly access information about the locations of schools across various boards and areas. Even if there are no schools in a particular board, the LEFT JOIN statement still includes all school boards in the report. |

1. **Conclusion section**

This individual project taught us important information on the distribution of public schools, such as which categories they fit into and who is in charge of them. To discover trends in the distribution of languages, educational standards, and the overall number of board-governed institutions, the study examined schools based on their type, location, and board affiliation. Because database queries could show both densely populated and sparsely schooled areas, they provided a decision-making framework. The addition of views and school categorization to the database increased data access and comprehension. As a result, we can more accurately comprehend the dataset and derive relevant insights.